Assignment 4

Khutso Ledwaba

2022-10-30

```
library("ggplot2")
library("gmodels")
library("caret")
library("ISLR")
library("class")
library("e1071")
library("dplyr")
library("melt")
library("reshape")
library("reshape2")
library("readr")
library("ISLR")
library("pROC")
library("cluster")

#Calling of the data file

Pharma <- read.csv(file = 'Pharmaceuticals.csv')
Pharma
```

```
##    Symbol                                Name Market_Cap Beta PE_Ratio  ROE  ROA
## 1     ABT               Abbott Laboratories      68.44 0.32     24.7 26.4 11.8
## 2     AGN                    Allergan, Inc.       7.58 0.41     82.5 12.9  5.5
## 3     AHM                      Amersham plc       6.30 0.46     20.7 14.9  7.8
## 4     AZN                   AstraZeneca PLC      67.63 0.52     21.5 27.4 15.4
## 5     AVE                          Aventis      47.16 0.32     20.1 21.8  7.5
## 6     BAY                         Bayer AG      16.90 1.11     27.9  3.9  1.4
## 7     BMY      Bristol-Myers Squibb Company      51.33 0.50     13.9 34.8 15.1
## 8    CHTT                      Chattem, Inc       0.41 0.85     26.0 24.1  4.3
## 9     ELN             Elan Corporation, plc       0.78 1.08      3.6 15.1  5.1
## 10    LLY             Eli Lilly and Company      73.84 0.18     27.9 31.0 13.5
## 11    GSK               GlaxoSmithKline plc     122.11 0.35     18.0 62.9 20.3
## 12    IVX                  IVAX Corporation       2.60 0.65     19.9 21.4  6.8
## 13    JNJ                Johnson & Johnson     173.93 0.46     28.4 28.6 16.3
## 14    MRX Medicis Pharmaceutical Corporation       1.20 0.75     28.6 11.2  5.4
## 15    MRK                  Merck & Co., Inc.     132.56 0.46     18.9 40.6 15.0
## 16    NVS                        Novartis AG      96.65 0.19     21.6 17.9 11.2
## 17    PFE                         Pfizer Inc     199.47 0.65     23.6 45.6 19.2
## 18    PHA             Pharmacia Corporation      56.24 0.40     56.5 13.5  5.7
## 19    SGP       Schering-Plough Corporation      34.10 0.51     18.9 22.6 13.3
## 20    WPI      Watson Pharmaceuticals, Inc.       3.26 0.24     18.4 10.2  6.8
## 21    WYE                             Wyeth      48.19 0.63     13.1 54.9 13.4
```

```
##    Asset_Turnover Leverage Rev_Growth Net_Profit_Margin Median_Recommendation
## 1             0.7     0.42       7.54              16.1          Moderate Buy
## 2             0.9     0.60       9.16               5.5          Moderate Buy
## 3             0.9     0.27       7.05              11.2            Strong Buy
## 4             0.9     0.00      15.00              18.0         Moderate Sell
## 5             0.6     0.34      26.81              12.9          Moderate Buy
## 6             0.6     0.00      -3.17               2.6                 Hold
## 7             0.9     0.57       2.70              20.6         Moderate Sell
## 8             0.6     3.51       6.38               7.5          Moderate Buy
## 9             0.3     1.07      34.21              13.3         Moderate Sell
## 10            0.6     0.53       6.21              23.4                 Hold
## 11            1.0     0.34      21.87              21.1                 Hold
## 12            0.6     1.45      13.99              11.0                 Hold
## 13            0.9     0.10       9.37              17.9          Moderate Buy
## 14            0.3     0.93      30.37              21.3          Moderate Buy
## 15            1.1     0.28      17.35              14.1                 Hold
## 16            0.5     0.06      -2.69              22.4                 Hold
## 17            0.8     0.16      25.54              25.2          Moderate Buy
## 18            0.6     0.35      15.00               7.3                 Hold
## 19            0.8     0.00       8.56              17.6                 Hold
## 20            0.5     0.20      29.18              15.1         Moderate Sell
## 21            0.6     1.12       0.36              25.5                 Hold
##       Location Exchange
## 1           US     NYSE
## 2       CANADA     NYSE
## 3           UK     NYSE
## 4           UK     NYSE
## 5       FRANCE     NYSE
## 6      GERMANY     NYSE
## 7           US     NYSE
## 8           US   NASDAQ
## 9      IRELAND     NYSE
## 10          US     NYSE
## 11          UK     NYSE
## 12          US     AMEX
## 13          US     NYSE
## 14          US     NYSE
## 15          US     NYSE
## 16 SWITZERLAND     NYSE
## 17          US     NYSE
## 18          US     NYSE
## 19          US     NYSE
## 20          US     NYSE
## 21          US     NYSE
```

```
str(Pharma)
```

```
## 'data.frame':    21 obs. of  14 variables:
##  $ Symbol               : chr  "ABT" "AGN" "AHM" "AZN" ...
##  $ Name                 : chr  "Abbott Laboratories" "Allergan, Inc." "Amersham plc" "AstraZeneca PL
##  $ Market_Cap           : num  68.44 7.58 6.3 67.63 47.16 ...
##  $ Beta                 : num  0.32 0.41 0.46 0.52 0.32 1.11 0.5 0.85 1.08 0.18 ...
##  $ PE_Ratio             : num  24.7 82.5 20.7 21.5 20.1 27.9 13.9 26 3.6 27.9 ...
##  $ ROE                  : num  26.4 12.9 14.9 27.4 21.8 3.9 34.8 24.1 15.1 31 ...
```

```
##  $ ROA                 : num  11.8 5.5 7.8 15.4 7.5 1.4 15.1 4.3 5.1 13.5 ...
##  $ Asset_Turnover      : num  0.7 0.9 0.9 0.9 0.6 0.6 0.9 0.6 0.3 0.6 ...
##  $ Leverage            : num  0.42 0.6 0.27 0 0.34 0 0.57 3.51 1.07 0.53 ...
##  $ Rev_Growth          : num  7.54 9.16 7.05 15 26.81 ...
##  $ Net_Profit_Margin   : num  16.1 5.5 11.2 18 12.9 2.6 20.6 7.5 13.3 23.4 ...
##  $ Median_Recommendation: chr  "Moderate Buy" "Moderate Buy" "Strong Buy" "Moderate Sell" ...
##  $ Location            : chr  "US" "CANADA" "UK" "UK" ...
##  $ Exchange            : chr  "NYSE" "NYSE" "NYSE" "NYSE" ...
```

```
head(Pharma)
```

```
##   Symbol                Name Market_Cap Beta PE_Ratio  ROE  ROA Asset_Turnover
## 1    ABT Abbott Laboratories      68.44 0.32     24.7 26.4 11.8            0.7
## 2    AGN      Allergan, Inc.       7.58 0.41     82.5 12.9  5.5            0.9
## 3    AHM        Amersham plc       6.30 0.46     20.7 14.9  7.8            0.9
## 4    AZN     AstraZeneca PLC      67.63 0.52     21.5 27.4 15.4            0.9
## 5    AVE             Aventis      47.16 0.32     20.1 21.8  7.5            0.6
## 6    BAY            Bayer AG      16.90 1.11     27.9  3.9  1.4            0.6
##   Leverage Rev_Growth Net_Profit_Margin Median_Recommendation Location Exchange
## 1     0.42       7.54              16.1          Moderate Buy       US     NYSE
## 2     0.60       9.16               5.5          Moderate Buy   CANADA     NYSE
## 3     0.27       7.05              11.2            Strong Buy       UK     NYSE
## 4     0.00      15.00              18.0         Moderate Sell       UK     NYSE
## 5     0.34      26.81              12.9          Moderate Buy   FRANCE     NYSE
## 6     0.00      -3.17               2.6                  Hold  GERMANY     NYSE
```

##Question 1

```
Pharma<- Pharma[,3:11]
#Display the first rows in the data
head(Pharma)
```

```
##   Market_Cap Beta PE_Ratio  ROE  ROA Asset_Turnover Leverage Rev_Growth
## 1      68.44 0.32     24.7 26.4 11.8            0.7     0.42       7.54
## 2       7.58 0.41     82.5 12.9  5.5            0.9     0.60       9.16
## 3       6.30 0.46     20.7 14.9  7.8            0.9     0.27       7.05
## 4      67.63 0.52     21.5 27.4 15.4            0.9     0.00      15.00
## 5      47.16 0.32     20.1 21.8  7.5            0.6     0.34      26.81
## 6      16.90 1.11     27.9  3.9  1.4            0.6     0.00      -3.17
##   Net_Profit_Margin
## 1              16.1
## 2               5.5
## 3              11.2
## 4              18.0
## 5              12.9
## 6               2.6
```

```
#Scale all the  ariables in the data frame.
Pharma <- scale(Pharma)
head(Pharma)
```

```
##      Market_Cap        Beta    PE_Ratio         ROE        ROA Asset_Turnover
```

```
## [1,]  0.1840960 -0.80125356 -0.04671323  0.04009035  0.2416121      0.0000000
## [2,] -0.8544181 -0.45070513  3.49706911 -0.85483986 -0.9422871      0.9225312
## [3,] -0.8762600 -0.25595600 -0.29195768 -0.72225761 -0.5100700      0.9225312
## [4,]  0.1702742 -0.02225704 -0.24290879  0.10638147  0.9181259      0.9225312
## [5,] -0.1790256 -0.80125356 -0.32874435 -0.26484883 -0.5664461     -0.4612656
## [6,] -0.6953818  2.27578267  0.14948233 -1.45146000 -1.7127612     -0.4612656
##        Leverage Rev_Growth Net_Profit_Margin
## [1,] -0.2120979 -0.5277675        0.06168225
## [2,]  0.0182843 -0.3811391       -1.55366706
## [3,] -0.4040831 -0.5721181       -0.68503583
## [4,] -0.7496565  0.1474473        0.35122600
## [5,] -0.3144900  1.2163867       -0.42597037
## [6,] -0.7496565 -1.4971443       -1.99560225
```

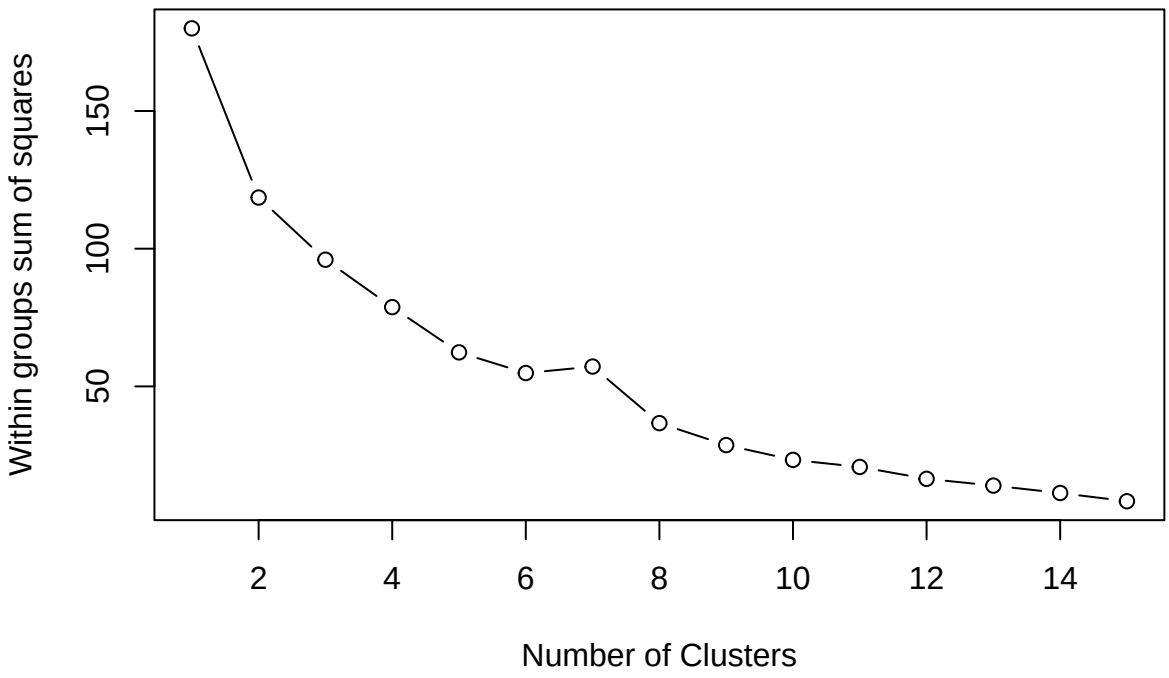
```
#The number of clusters required for the analysis
Clusters <- (nrow(Pharma)-1)*sum(apply(Pharma,2,var))
Clusters
```

```
## [1] 180
```

```
for (i in 2:15) Clusters[i] <- sum(kmeans(Pharma,centers=i)$withinss)
Clusters
```

```
##  [1] 180.00000 118.56934  95.99420  78.77959  62.35443  54.84345  57.19428
##  [8]  36.65309  28.67945  23.31081  20.74157  16.42592  13.97318  11.28999
## [15]   8.29565
```

```
# Graph to analyze the clusters

plot(1:15, Clusters, type="b", xlab="Number of Clusters",ylab="Within groups sum of squares")
```



Question 2 interpretation of the clusters

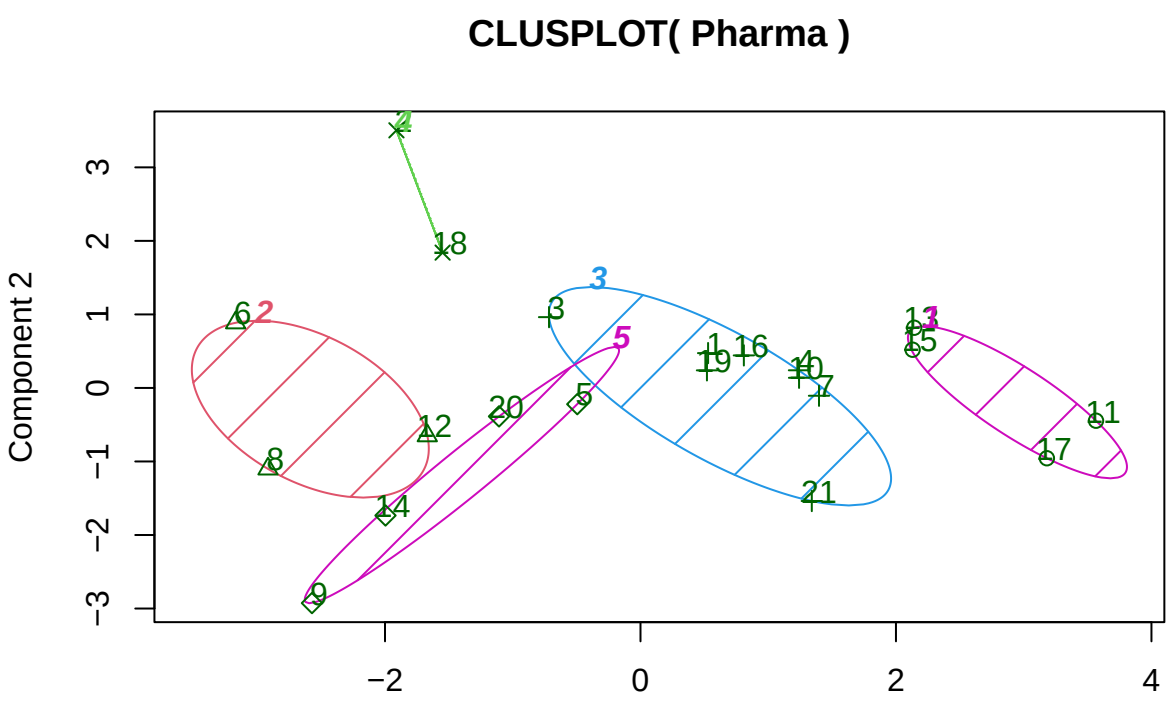
```
#Partitioning the data

 fit <- kmeans(Pharma, 5)
 aggregate(Pharma,by=list(fit$cluster),FUN=mean)
```

```
##   Group.1  Market_Cap       Beta    PE_Ratio        ROE        ROA
## 1       1  1.69558112 -0.1780563 -0.19845823  1.2349879  1.3503431
## 2       2 -0.87051511  1.3409869 -0.05284434 -0.6184015 -1.1928478
## 3       3 -0.03142211 -0.4360989 -0.31724852  0.1950459  0.4083915
## 4       4 -0.43925134 -0.4701800  2.70002464 -0.8349525 -0.9234951
## 5       5 -0.76022489  0.2796041 -0.47742380 -0.7438022 -0.8107428
##   Asset_Turnover    Leverage Rev_Growth Net_Profit_Margin
## 1      1.1531640 -0.46807818  0.4671788       0.591242521
## 2     -0.4612656  1.36644699 -0.6912914      -1.320000179
## 3      0.1729746 -0.27449312 -0.7041516       0.556954446
## 4      0.2306328 -0.14170336 -0.1168459      -1.416514761
## 5     -1.2684804  0.06308085  1.5180158      -0.006893899
```

```
Pharma1 <- data.frame(Pharma, fit$cluster)
Pharma1
```

```
##    Market_Cap        Beta    PE_Ratio         ROE        ROA Asset_Turnover
## 1   0.1840960 -0.80125356 -0.04671323  0.04009035  0.2416121      0.0000000
## 2  -0.8544181 -0.45070513  3.49706911 -0.85483986 -0.9422871      0.9225312
## 3  -0.8762600 -0.25595600 -0.29195768 -0.72225761 -0.5100700      0.9225312
## 4   0.1702742 -0.02225704 -0.24290879  0.10638147  0.9181259      0.9225312
## 5  -0.1790256 -0.80125356 -0.32874435 -0.26484883 -0.5664461     -0.4612656
## 6  -0.6953818  2.27578267  0.14948233 -1.45146000 -1.7127612     -0.4612656
## 7  -0.1078688 -0.10015669 -0.70887325  0.59693581  0.8617498      0.9225312
## 8  -0.9767669  1.26308721  0.03299122 -0.11237924 -1.1677918     -0.4612656
## 9  -0.9704532  2.15893320 -1.34037772 -0.70899938 -1.0174553     -1.8450624
## 10  0.2762415 -1.34655112  0.14948233  0.34502953  0.5610770     -0.4612656
## 11  1.0999201 -0.68440408 -0.45749769  2.45971647  1.8389364      1.3837968
## 12 -0.9393967  0.48409069 -0.34100657 -0.29136529 -0.6979905     -0.4612656
## 13  1.9841758 -0.25595600  0.18013789  0.18593083  1.0872544      0.9225312
## 14 -0.9632863  0.87358895  0.19240011 -0.96753478 -0.9610792     -1.8450624
## 15  1.2782387 -0.25595600 -0.40231769  0.98142435  0.8429577      1.8450624
## 16  0.6654710 -1.30760129 -0.23677768 -0.52338423  0.1288598     -0.9225312
## 17  2.4199899  0.48409069 -0.11415545  1.31287998  1.6322239      0.4612656
## 18 -0.0240846 -0.48965495  1.90298017 -0.81506519 -0.9047030     -0.4612656
## 19 -0.4018812 -0.06120687 -0.40231769 -0.21181593  0.5234929      0.4612656
## 20 -0.9281345 -1.11285216 -0.43297324 -1.03382590 -0.6979905     -0.9225312
## 21 -0.1614497  0.40619104 -0.75792214  1.92938746  0.5422849     -0.4612656
##       Leverage  Rev_Growth Net_Profit_Margin fit.cluster
## 1  -0.21209793 -0.52776752        0.06168225           3
## 2   0.01828430 -0.38113909       -1.55366706           4
## 3  -0.40408312 -0.57211809       -0.68503583           3
## 4  -0.74965647  0.14744734        0.35122600           3
## 5  -0.31449003  1.21638667       -0.42597037           5
## 6  -0.74965647 -1.49714434       -1.99560225           2
## 7  -0.02011273 -0.96584257        0.74744375           3
## 8   3.74279705 -0.63276071       -1.24888417           2
## 9   0.61983791  1.88617085       -0.36501379           5
## 10 -0.07130879 -0.64814764        1.17413980           3
## 11 -0.31449003  0.76926048        0.82363947           1
## 12  1.10620040  0.05603085       -0.71551412           2
## 13 -0.62166634 -0.36213170        0.33598685           1
## 14  0.44065173  1.53860717        0.85411776           5
## 15 -0.39128411  0.36014907       -0.24310064           1
## 16 -0.67286239 -1.45369888        1.02174835           3
## 17 -0.54487226  1.10143723        1.44844440           1
## 18 -0.30169102  0.14744734       -1.27936246           4
## 19 -0.74965647 -0.43544591        0.29026942           3
## 20 -0.49367621  1.43089863       -0.09070919           5
## 21  0.68383297 -1.17763919        1.49416183           3
```

```
clusplot(Pharma, fit$cluster, color=TRUE, shade=TRUE, labels=2, lines=0)
```




```
# From the output we see that Cluster 1 has the lowest Leverage.
# Cluster 2 has the lowest ROE, lowest ROA,lowest Net Profit Margin, highest PE ratio,  and lowest Asse
# Cluster 3 has the highest ROA, highest Asset Turnover Market Cap, highest ROE
# Cluster 4 has the lowest Beta, lowest PE Ratio, Net Profit Margin and lowest Rev growth.
# Cluster 5 has the highest Leverage, highest Rev growth, highest Beta and lowest Market Cap.
```

Question 3 patterns identified

```
# Cluster 3 has been identified to have buy media recommendations as it is listed with highest Market C
# Cluster 4 which has the highest Net Profit Margin,lowest Rev growth, lowest Beta and lowest PE Ratio
```

Question 4 naming of the clusters

```
# Cluster 1 - Lowest Leverage cluster
# Cluster 2 - Net Profit Margin cluster (lowest ROE, lowest ROA)
# Cluster 3 - Asset Turnover cluster (highest ROA, highest Asset Turnover Market Cap)
# Cluster 4 - Revenue growth cluster (lowest Beta, lowest PE Ratio, Net Profit Margin)
# Cluster 5 - Low Market Cap cluster (highest Leverage, highest Rev growth, highest Beta)
```